Guesswork and Entropy

David Malone and Wayne G. Sullivan

***Abstract*—We derive the moments of the *guesswork*, the number of attempts required to correctly guess the output of a random source, for a source determined by a Markov chain via a large deviations type estimate. These moments are related to the Perron–Frobenius eigenvalue of the matrix formed by element-wise powers of the Markov chain's transition matrix.**

***Index Terms*—Guesswork, large deviation, Markov chain, Rényi entropy.**

I. Introduction

CONSIDER the problem of guessing the value of a discrete random variable. Can we quantify the difficulty involved in guessing it? Metrics considered in recent years include the mean and the moments of the number of guesses required to get the correct answer. For example, in [1] the mean number of guesses is compared to the Shannon entropy and in [2] it is compared to the Rényi entropy.

The guessing strategy we consider is the usual optimal one, where symbols are guessed in decreasing order of probability. If the symbols produced by the source are relabeled so that p_1 is the most likely and the sequence p_i is nonincreasing then the expected number of guesses is

$$G(p) = \sum_i i p_i. \tag{1}$$

In [3] this is referred to as the *guesswork*. Note, that this may not always be a good measure of "guessability" depending on the application, and [3] also considers other metrics.

In the context of repeated experiments some care is needed in the computation of (1). When calculating expectations for n independent and identically distributed (i.i.d.) sources, we look at sums of the form

$$\sum_{n_1,\ldots,n_r} \binom{n}{n_1,\ldots,n_r} p_1^{n_1},\ldots,p_r^{n_r} f(p). \tag{2}$$

If the function $f(p)$ is relatively small, then the most important term in this sum is the one which maximizes the product of the multinomial coefficient and the probabilities. This term will have $n_k/n \approx p_k$. These points correspond to the typical set of asymptotic equipartition.

When calculating guesswork, $f(p) = \operatorname{rank}(p)$ and the sum we consider is closer to

$$\sum_{n_1,\ldots,n_r} \binom{n}{n_1,\ldots,n_r}^2 p_1^{n_1},\ldots,p_r^{n_r}. \tag{3}$$

Manuscript received April 12, 2002; revised August 26, 2003.

The authors are with the Communications Network Research Institute, Dublin Institute of Technology, Dublin, Ireland.

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2004.824921

Here the largest terms will be those with $n_k/n \approx c\sqrt{p_k}$, where c is a normalizing constant. Thus, the dominant terms for the guesswork problem are different from those for the coding problem. In [4], Arikan employs clever inequalities to produce estimates of the guesswork. Our method is to apply direct calculations to extend this result to Markov chains.

Let us now state precisely the problem we consider. For the probability distribution $\{p_1,\ldots,p_m\}$, $p_k \geq p_{k+1}$, and $\alpha > 0$, the αth guesswork moment G^α is given by

$$G^\alpha := \sum_{i=1}^{m} i^\alpha p_i. \tag{4}$$

Let $\mathtt{A} = \{1,\ldots,r\}$ be a finite alphabet with $r > 1$ characters. Let P be a stationary distribution on $\mathtt{A}^{\boldsymbol{N}}$, with P_n denoting the restrictions of P to $\mathtt{A}^n$. We seek

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) \tag{5}$$

where $\lg$ denotes the logarithm to any base > 1. Arikan [4] has shown that in the independent case, where P_n is the product of $p_1,\ldots,p_r$

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) = (1+\alpha)\lg \sum_{i=1}^{m} p_i^{1/(1+\alpha)}. \tag{6}$$

A. Notation

It is convenient to specify the notation we use for the irreducible (possibly periodic) Markov chain P on $\mathtt{A}^{\boldsymbol{N}}$. The restriction of P to $\mathtt{A}^n$ is denoted P_n. We assume P has the stochastic matrix $U = (U_{ab})$ and invariant probability (u_a) so that for $\omega \in \mathtt{A}^{n+1}$

$$P_{n+1}(\omega) = u_{\omega_1} \prod_{i=1}^{n} U_{\omega_i \omega_{i+1}}. \tag{7}$$

Theorem I.1: Let P be the irreducible Markov chain specified above. Then for $\alpha > 0$

$$\lim_n \frac{1}{n} \lg G^\alpha(P_n) = (1+\alpha)\lg \lambda \tag{8}$$

where λ is the Perron–Frobenius eigenvalue of the matrix with entries $U_{ab}^{1/(1+\alpha)}$.

II. Proofs

We outline our original proof and give a shorter proof suggested by a referee. Both proofs use the Perron–Frobenius theorem (see [5]):

Theorem II.1: Let V be an irreducible matrix on $\mathtt{A} \times \mathtt{A}$ with nonnegative entries. Then V has left and right eigenvectors

$(v_a : a \in \mathtt{A})$, $(w_a : a \in \mathtt{A})$ all of whose entries are strictly positive. The corresponding eigenvalue λ is real and has the property that if λ' is any other real or complex eigenvalue of V, then $|\lambda'| \le \lambda$.

The basic steps of our original proof are as follows. We assigned the probability $P_{n+1}(\omega)$ given by (7) to the sample path ω. Let

$$n_{ab}(n,\omega) := |\{i : 1 \le i \le n,\ \omega_i = a,\ \omega_{i+1} = b\}| \quad (9)$$

$$n_a := \sum_{b \in \mathrm{A}} n_{ab} \quad (10)$$

with $|\cdot|$ denoting cardinality. Then

$$P_{n+1}(\omega) = u_{\omega_1} \prod_{a,b \in \mathrm{A}^2} U_{ab}^{n_{ab}(\omega)}. \quad (11)$$

Given $(\widetilde{n}_{ab})$ corresponding to some path $\widetilde{\omega}$, we defined

$$e\,(n, c, (\widetilde{n}_{ab})) := \left|\{\omega \in \mathtt{A}^{n+1} : \omega_1 = c,\ n_{ab}(n,\omega) = \widetilde{n}_{ab}\ \forall\, ab \in \mathtt{A}^2\}\right|. \quad (12)$$

We then deduced

$$\lim_n \lg G^\alpha(P_{n+1})/(n+1) = \lim_n \frac{1}{n+1} \lg \max_\omega P_{n+1}(\omega) \Big(e\,(n, c, (n_{ab}(\omega)))\Big)^{1+\alpha}. \quad (13)$$

In the case that $\alpha = 0$, the maximum occurs for paths ω which are typical in the sense of asymptotic equipartition for the distribution associated with U. Define V with $V_{ab} := U_{ab}^{1/1+\alpha}$ and let λ, $\{v_a\}$, $\{w_b\}$ be the Perron–Frobenius eigenvalue and eigenvectors of V, so that

$$\sum_a v_a = 1,\ \sum_a v_a V_{ab} = \lambda v_b,\ \sum_b w_b V_{ab} = \lambda w_a. \quad (14)$$

The paths which maximize

$$\Big(P_{n+1}^{1/(1+\alpha)}(\omega)\, e\,(n, c, (n_{ab}(\omega)))\Big)^{1+\alpha} \quad (15)$$

correspond to the typical paths of the Markov chain associated with the stochastic matrix with entries $V_{ab}w_b/\lambda w_a$. Our original proof consisted of deducing bounds for $e\,(n, c, (n_{ab}(\omega)))$ and using these bounds to compute (13). Here is a shorter proof of the theorem.

Proof: We start from the inequalities (see [4])

$$\frac{\|P_{n+1}\|_{1/(1+\alpha)}}{(1 + \ln |\mathtt{A}|^{n+1})^\alpha} \le G^\alpha(P_{n+1}) \le \|P_{n+1}\|_{1/(1+\alpha)} \quad (16)$$

where

$$\|P_{n+1}\|_{1/(1+\alpha)} := \left(\sum_{\omega \in \mathrm{A}^{n+1}} P_{n+1}^{1/(1+\alpha)}(\omega) \right)^{1+\alpha}. \quad (17)$$

Since

$$\lim_n \frac{\alpha}{n+1} \lg(1 + (n+1)\ln |\mathtt{A}|) = 0 \quad (18)$$

we deduce

$$\limsup_n \frac{\lg \|P_{n+1}\|_{1/(1+\alpha)}}{n+1} = \limsup_n \frac{\lg G^\alpha(P_{n+1})}{n+1} \quad (19)$$

and the same holds for $\liminf$. Thus, it suffices to show that $\lim_n \lg \|P_{n+1}\|_{1/(1+\alpha)}/(n+1)$ exists and has the required value. Now

$$P_{n+1}^{1/(1+\alpha)}(\omega) = \frac{u_{\omega_1}^{1/(1+\alpha)}}{v_{\omega_1}}\, v_{\omega_1} \prod_{a,b \in \mathrm{A}^2} V_{ab}^{n_{ab}}. \quad (20)$$

Matrix multiplication yields

$$\sum_{\omega \in \mathrm{A}^{n+1}} v_{\omega_1} \prod_{a,b \in \mathrm{A}^2} V_{ab}^{n_{ab}(\omega)} = \lambda^n. \quad (21)$$

Hence,

$$\min_{a \in \mathrm{A}} \frac{u_a}{v_a^{1+\alpha}} \le \frac{\|P_{n+1}\|_{1/(1+\alpha)}}{\lambda^{n(1+\alpha)}} \le \max_{a \in \mathrm{A}} \frac{u_a}{v_a^{1+\alpha}} \quad (22)$$

from which we deduce

$$\lim_n \frac{1}{n+1} \lg \|P_{n+1}\|_{1/(1+\alpha)} = (1+\alpha) \lg \lambda. \quad (23)$$

This completes the proof of Theorem I.1. □

Note that the above proof does not give any indication of the dominant term corresponding to (13).

A special case is that for which all the rows of the stochastic matrix $U = (U_{ab})$ are equal, which corresponds to independence. This yields (6).

REFERENCES

[1] J. L. Massey, "Guessing and entropy," in *Proc. IEEE Int. Symp. Information Theory*, 1994, p. 204.

[2] S. S. Dragomir and S. Boztas, "Two sided bounds on guessing moments," Dept. Math., Roy. Melbourne Inst. Technol., Melbourne, Australia, Res. Rep. 8, 1997.

[3] J. O. Pliam. (1999) The Disparity Between Work and Entropy in Cryptology. [Online]. Available: http://philby.ucsd.edu/cryptolib/1998/98-24.html

[4] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 99–105, Jan. 1996.

[5] D. Lind and B. Marcus, *Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995.